Protein Structure and Function by Comparative Model Building

JONATHAN GREER

Physical Biochemistry Laboratory
Computer-Assisted Molecular Design
Abbott Laboratories
Abbott Park, North Chicago, Illinois 60064

INTRODUCTION

Comparative modeling methods have been used to extend the experimentally determined three-dimensional structures of proteins to new molecules whose structure is closely related. Such techniques have been applied to derive model structures for α-lactalbumin,[1] α-lytic protease,[2] *Streptomyces* trypsin-like protein,[3] Ca^{++} binding proteins,[4] haptoglobin,[5] serine proteases,[6] including blood clotting factor Xa[7] and very recently renin,[8] a member of the acid protease family. Thus, comparative model building has been widely used for producing tentative structures of biologically interesting and important molecules.

In this study, we employ comparative modeling methods to begin exploring the nature of specificity between enzymes and their particular substrates. Initial structures for several serine proteases are derived.[5-7] Suitable peptides from the known macromolecular substrates of these enzymes are also modeled onto the enzyme active site[7] and their properties in relation to their respective enzymes are analyzed and compared. The ultimate goal is to gain a deeper and more detailed understanding of the molecular basis of enzyme-substrate and protein-ligand recognition and specificity.

METHODS

Comparative Modeling Method

A detailed description of the comparative modeling methods for the serine proteases used in this work has been published.[6] Briefly, by comparing the experimentally known structures of chymotrypsin,[9,10] trypsin,[11,12] and elastase,[13,14] the three-dimensional structures of the serine proteases can be parsed into structurally conserved regions (SCRs) and variable regions (VRs) (see FIG. 1). Sequence homology among these proteins lies almost exclusively in the SCRs (see Figure 2 of ref. 6).

To model a "new" serine protease, the sequence is aligned using the strong

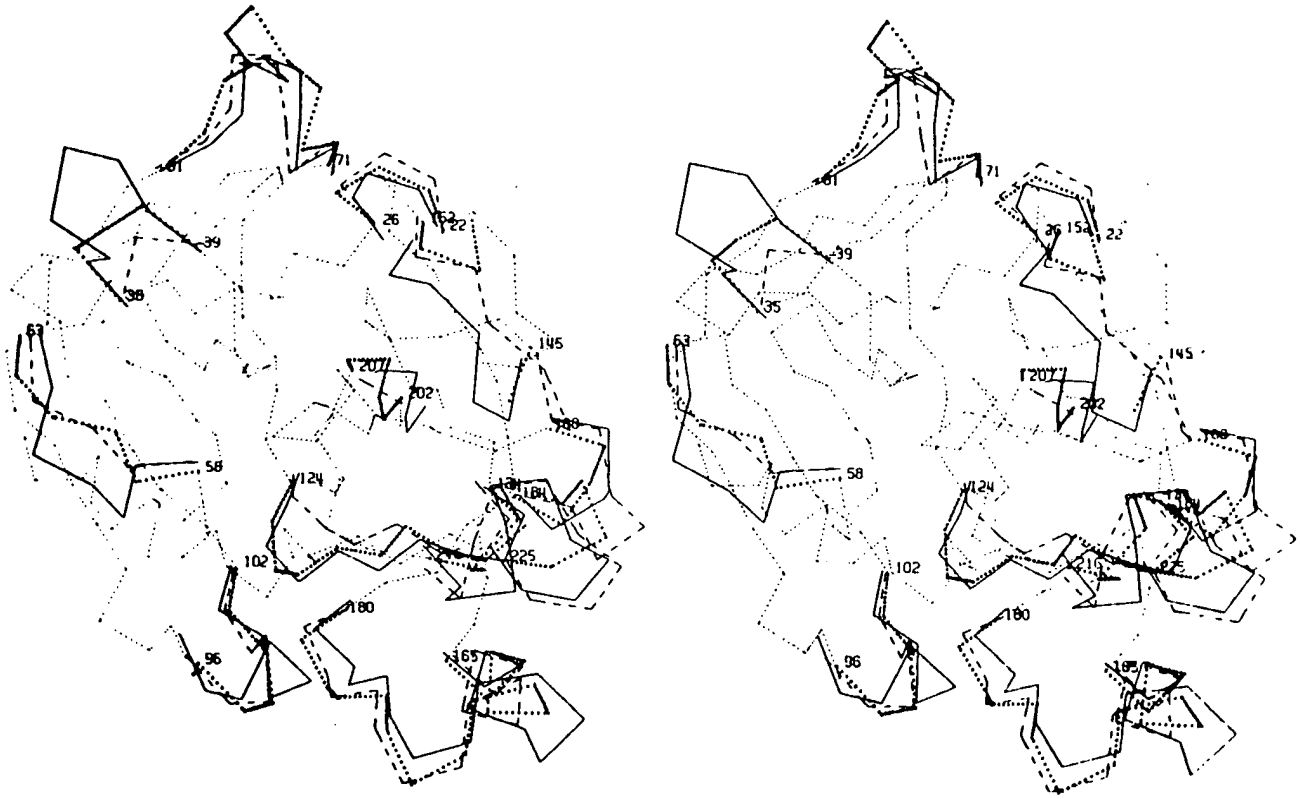

FIGURE 1. An α-carbon plot of the structurally conserved regions (*dotted lines*) of chymotrypsin together with the different variable regions for each of the three serine protease structures: chymotrypsin (*dotted lines*), trypsin (*dashed lines*), and elastase (*solid lines*). The three conformations at a variable region are considered when modeling the structure of a "new" protein at that loop.

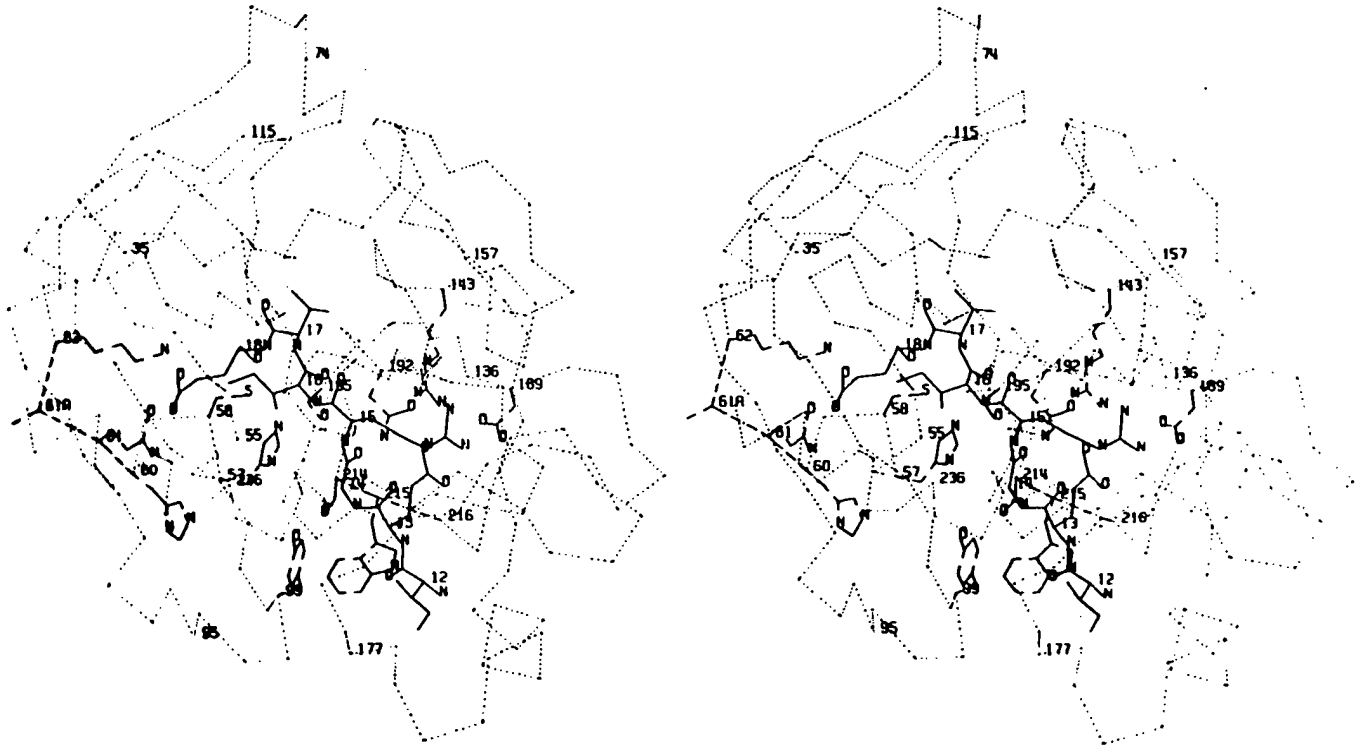

FIGURE 2. Model structure for the factor Xa-prothrombin complex in stereo. The α-carbon plot of factor Xa is shown (*dotted lines*) with the side chains (*dashed lines*) that are interacting with the prothrombin substrate (*solid lines*). Residue Arg P_1 (labeled 15) forms a salt bridge with Asp 189 of factor Xa. Glu P_3 (labeled 13) forms a bridge with Arg 143 and Glu P_3' (labeled 18) forms a bridge with Lys 62. Any larger side chain than Gly at site P_2 (labeled 14) would collide with Tyr 99 of factor Xa.

homology in the SCRs. The coordinates for the main chain in the SCRs are taken from any one of the known structures (since they are almost identical, see Figure 1 of ref. 6). The side chains are "mutated" to fit the sequence of the new protein.

Modeling the VRs is much more challenging. For each VR, the various conformations found amongst the known structures (FIG. 1) are examined as to length (see Table 3 of ref. 6) and residue character. If one of the known conformations fits the new sequence, its main chain coordinates are used directly with suitable replacement of side chains. Otherwise, modeling using energetics is necessary to achieve a reasonable tentative conformation for the respective VR. Modeling studies on eight serine protease sequences[6,7] show that more than 50% of the VRs can be modeled directly from the known structures whereas less than 5% fall in a class of large additions where modeling is not practical in the near future.

Modeling the Serine Proteases and Their Substrates

TABLE 1 lists the substrates and enzymes used in this study. The two serine proteases, blood clotting factors Xa and IXa, were sequenced by Titani *et al.*[15] and Katayama *et al.*,[16] respectively. The sequence alignments used for these two proteins are given in Figure 3 of ref. 6. The SCRs were modeled from elastase and the VRs as shown in TABLE 2. In the case of factor Xa (see ref. 7) three loops cannot be built without further energy analysis; however, these three VRs are distant from the active site region and thus do not concern us in this study. Factor IXa is closely related but not identical to factor Xa in the size of its VRs as can be seen from Figure 3 and Table 3 of ref. 6 and from TABLE 2. In factor IXa, only two loops have no direct known structure as a model. One of them, the VR at 36–38, lies just on the border of the substrate binding pocket. The other is distant from the active site region, as it was in factor Xa.

The substrate peptides, consisting of residues P_4,..., P_1, P_1',..., P_3' (using the standard substrate nomenclature[17]), were modeled after the crystallographically determined conformations of bovine pancreatic trypsin inhibitor,[11] of soybean trypsin inhibitor,[18,19] and of di- and tripeptide chloromethyl ketones bound to γ-chymotrypsin[20,21] and to *Streptomyces griseus* protease B.[22] All these inhibitors have a common main chain conformation between residues P_3 and P_3' (see ref. 7 for a more complete discussion of this). The side chain positions for residues P_3 and P_3' as well as the position of residue

TABLE 1. Substrates and Enzymes Used

Substrate	Enzyme
prothrombin	factor Xa
factor X	factor IXa
trypsinogen	trypsin

TABLE 2. Model Structures for the VRs in Factors Xa and IXa

	Factor Xa		Factor IXa	
VR	Model	Residues built	Model	Residues built
23–25	elastase	–	elastase	–
36–38	chymotrypsin	33–40	deletion	(at 37)
59–62	elastase	–	elastase	–
72–80	deletion	(at 76)	elastase	–
97–101	trypsin	95–102	elastase	–
116	deletion	(at 116)	elastase	–
124–133	addition	(at 131)	addition	(at 131)
146–151	elastase	–	elastase	–
166–179	trypsin	164–180	trypsin	164–180
185–187	trypsin	184–188	trypsin	184–188
203–206	elastase	–	elastase	–
217–224	chymotrypsin	216–226	chymotrypsin	216–226

P_4 were determined by modeling the substrate on the respective enzyme as previously described.[7]

Surface Representations and Electrostatic Potential Maps

The solvent exclusion surfaces of the substrates and enzymes were calculated by the method of Connolly[23,24] using a program obtained from him.

The electrostatic potential maps were calculated as follows: a point unit positive charge was placed on the solvent-accessible surface and the electrostatic potential calculated from Coulomb's Law as

$$E_{el} = \sum_i \frac{q_i}{r_i}$$

where q_i is the charge on the ith atom and r_i is the distance from the point positive charge to this atom. The charges on the atoms were taken from Kollman and his co-workers.[25] A unit dielectric constant is used here because we are interested in the interactions between enzyme and substrate at the complex interface from which water should be excluded. The major conclusions of this work are not changed if a dielectric constant $= r_i$ is used as reported by many workers.[26–28]

RESULTS AND DISCUSSION

The Structures of the Enzyme-Substrate Complexes

The details of the proposed enzyme-substrate complex for factor Xa and prothrombin have previously been described[7] (FIG. 2). The cleaved peptide bond lies between residues P_1 and P_1' in prothrombin (see TABLE 3). Residue P_1 is an Arg and forms a salt bridge with Asp 189 of factor Xa. This interac-

TABLE 3. Sequence of Cleaved Peptide in Substrate

	P_4[a]	P_3	P_2	P_1	P_1'	P_2'	P_3'
Substrate	12[b]	13	14	15	16	17	18
Prothrombin	Ile	Glu	Gly	Arg	Ile	Val	Glu
Factor X	Gln	Val	Val	Arg	Ile	Val	Gly
Trypsinogen	Asp	Asp	Asp	Lys	Ile	Val	Gly

[a] Nomenclature as in ref. 17.

[b] This residue numbering corresponds to that for chymotrypsinogen which is often used for comparing serine protease sequences. These numbers are used to label the residues of the substrate peptides in FIGS. 2–9.

tion is typical of many serine proteases and is the molecular basis of the primary specificity shown by these enzymes, including trypsin, to cleave only after Lys or Arg residues.[29] The main chain of residues P_2 and P_3 forms an antiparallel β-sheet with residues 216–218 of the factor Xa enzyme. This feature appears to be common to all the inhibitor structures examined experimentally and is assumed in all the complexes modeled in this study. Residue P_2 is a Gly and thus has no side chain; any larger side chain at this point would collide with the phenolic hydroxyl of Tyr 99 on factor Xa. The glutamate side chains at P_3 and P_3' form specific salt bridges with Arg 143 and Lys 62 on factor Xa, respectively. Finally, Ile P_4, Ile P_1', and Val P_2' all lie in hydrophobic regions of the molecule. Thus, to sum up, in addition to the Arg-Asp salt bridge conferring the primary specificity, there are two other salt bridges, a stereogeometric requirement for a Gly, and numerous hydrophobic interactions.

The complex between blood clotting factor X as a substrate and its activating enzyme factor IXa was also examined in detail (FIG. 3). The primary specificity interaction is the same as above: Arg P_1 of factor X with Asp 189 of factor IXa. The side chains of Val P_2 and Val P_3 appear to make few close interactions, nor do they seem to lie in a particularly hydrophobic environment. On the other hand, Gln P_4 has the possibility of making several hydrogen bonds with the hydroxyl side chains of Thr 172 and Ser 175. On the other side of the cleaved bond, the relatively invariant Ile P_1' and Val P_2' lie in conserved hydrophobic pockets on factor IXa. Residue P_3' lies immediately adjacent to the VR at positions 36–38, which cannot be modeled in factor IXa without detailed energetic analysis since no experimentally known conformation exists for this sequence (see METHODS and TABLE 2). However, the occurrence of Gly P_3' in factor X avoids the need to model this VR accurately because it has no side chain to fit to the enzyme as there was in the prothrombin-factor Xa complex (see above and ref. 7 for a detailed discussion). Therefore, the major interactions between factor X and factor IXa, in addition to the Arg P_1-Asp 189 salt bridge, are two or three hydrogen bonds and several hydrophobic contacts.

The contrast between the two enzyme-substrate complexes described above is quite striking. Whereas the occurrence of two additional specific salt bridges

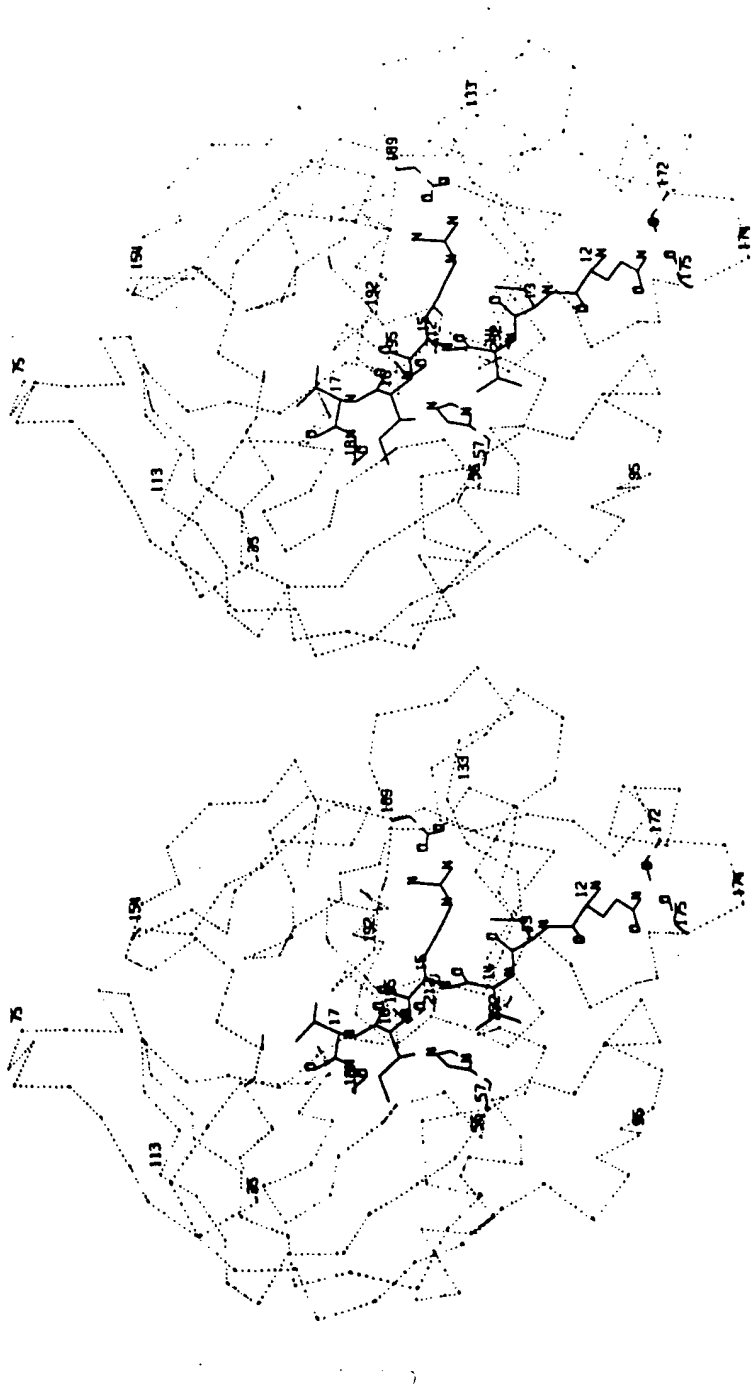

FIGURE 3. Model structure for the factor IXa-factor X complex in stereo. Factor IXa is presented as an α-carbon plot (*dotted lines*) with the relevant side chains (*dashed lines*). The substrate, factor X (*solid lines*), forms the primary specificity salt bridge between Arg P_1 (labeled 15) and Asp 189 of factor IXa. The only other polar interaction lies between Gln P_4 (labeled 12) and Thr 172 and Ser 175. There may be further interactions between the Gln and residues on the adjacent loop around position 99. The side chains of Val P_2 and Val P_3 (labeled 14 and 13, respectively) seem to make very few interactions.

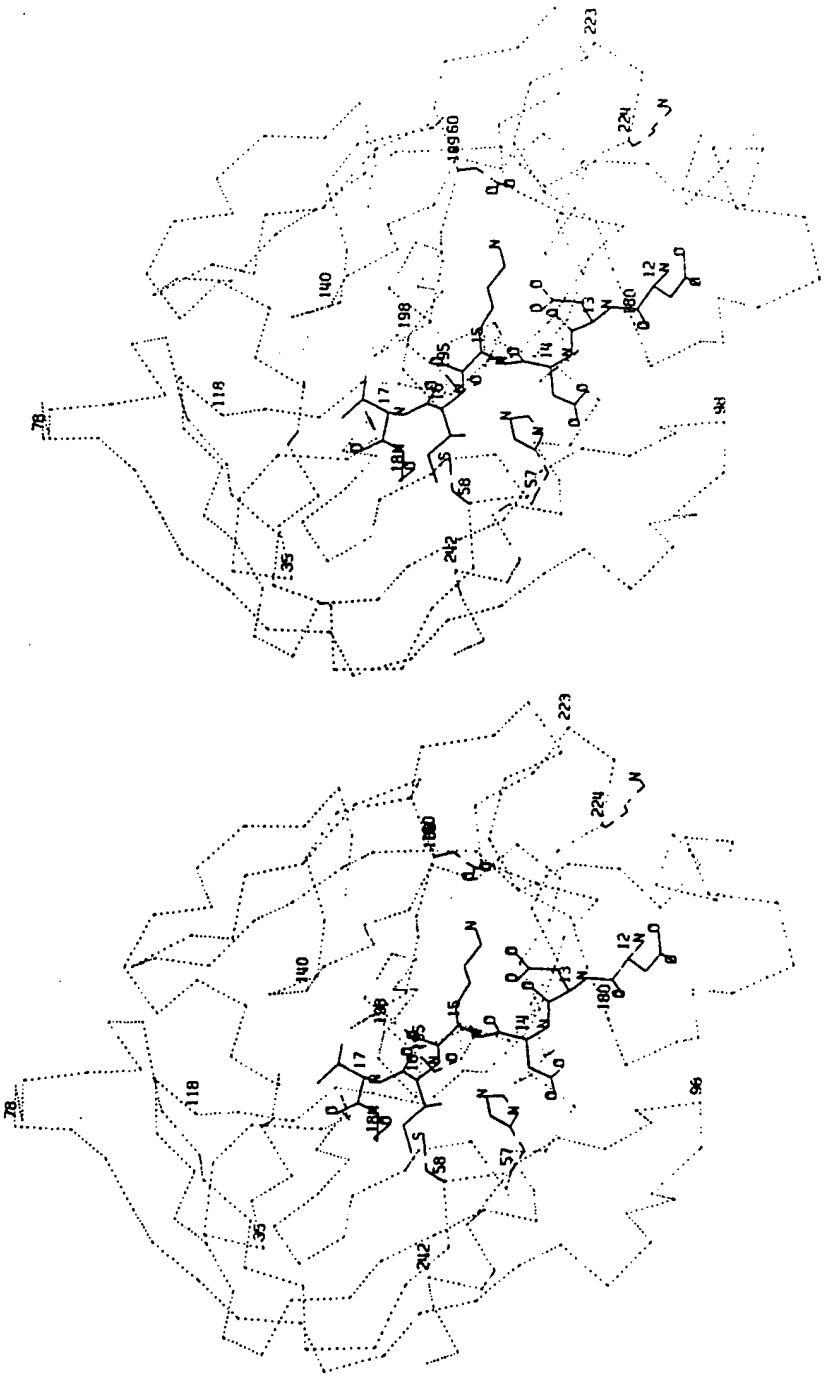

FIGURE 4. Model structure for the complex of the activation peptide of trypsinogen (*solid lines*) with the experimental structure of trypsin (*dotted and dashed lines*) shown in stereo. Lys P_1 (labeled 15) forms a salt bridge with Asp 189 of trypsin. Asp P_2 and P_3 (labeled 14 and 13) point out into solvent as does Asp P_4 (labeled 12). The latter may interact with Lys 224 of trypsin, shown here pointing into solvent also.

beyond the primary specificity interaction seems likely in factor Xa-prothrombin, no such additional charge-charge interactions occur in the factor IXa-factor X complex. In fact, the side chains of Val P_2, Val P_3', and of course, Gly P'_3 in factor X seem to contribute very little to either specificity or binding to the factor IXa enzyme.

The very different nature of the interactions found in the factor Xa-prothrombin and factor IXa-factor X complexes encouraged us to examine another enzyme-substrate interaction in order to see how different these substrates are from each other. Therefore, the trypsinogen activation sequence (TABLE 3) was examined in the appropriate conformation for binding to an activating serine protease, in this case taken to be trypsin (see FIG. 4 and TABLE 1). Once again, the primary specificity salt bridge appears, this time between Lys P_1 and Asp 189 of trypsin. The three residues prior to the Lys are aspartates. No countercharge occurs on trypsin for either Asp P_2 or Asp P_3. The ε-amino group of Lys 224 on trypsin is approximately 8 Å from Asp P_4, but can be rotated to lie in close association with that carboxylate. On the other side of Lys P_1, Ile P_1' and Val P_2' lie in their usual, conserved hydrophobic environments. Gly at P_3', of course, has no side chain to interact with the enzyme.

Overall, the trypsinogen peptide seems to interact even less with the trypsin active site region than does factor X with factor IXa. In that latter case, H-bonds were formed to Gln P_4 and valines P_2 and P_3 did provide some hydrophobic interactions beyond the Ile P_1', Val P_2' contacts. In the trypsin-trypsinogen complex, there is almost no interaction of side chains other than that of Ile P_1' and Val P_2'.

This result accords well with the fact that trypsin has lower secondary specificity than the other enzymes studied here. The absence of specific, limiting interactions between the enzyme and substrate side chains at positions P_2 through P_4 and P_3' should allow a wide variety of peptides to bind to the active site. Since trypsin is presumably intended to cleave denatured and partially fragmented proteins, any charged or polar side chains on the substrate which would not interact with the enzyme (such as Asp P_2 through P_4) could be satisfied by the buffer and the solvent.

Comparison of Substrate Properties

It is clear, from examining the three peptide sequences in TABLE 3 and FIGS. 2–4, that each substrate will make quite different hydrophilic and hydrophobic interactions with its respective cleaving enzyme. Consequently,

FIGURE 5. Electrostatic potential surface maps for the three substrate peptides of TABLE 1 shown in stereo. The electrostatic potential is represented as follows: < -15 kcal/mole, red; -15 to -5 kcal/mole, orange; -5 to 5 kcal/mole, green; 5 to 15 kcal/mole, light blue; >15 kcal/mole, blue. *Top* is prothrombin activation peptide, *middle* is factor X activation peptide, and *bottom* is trypsinogen activation peptide. The statistics for these maps are presented in TABLE 4.

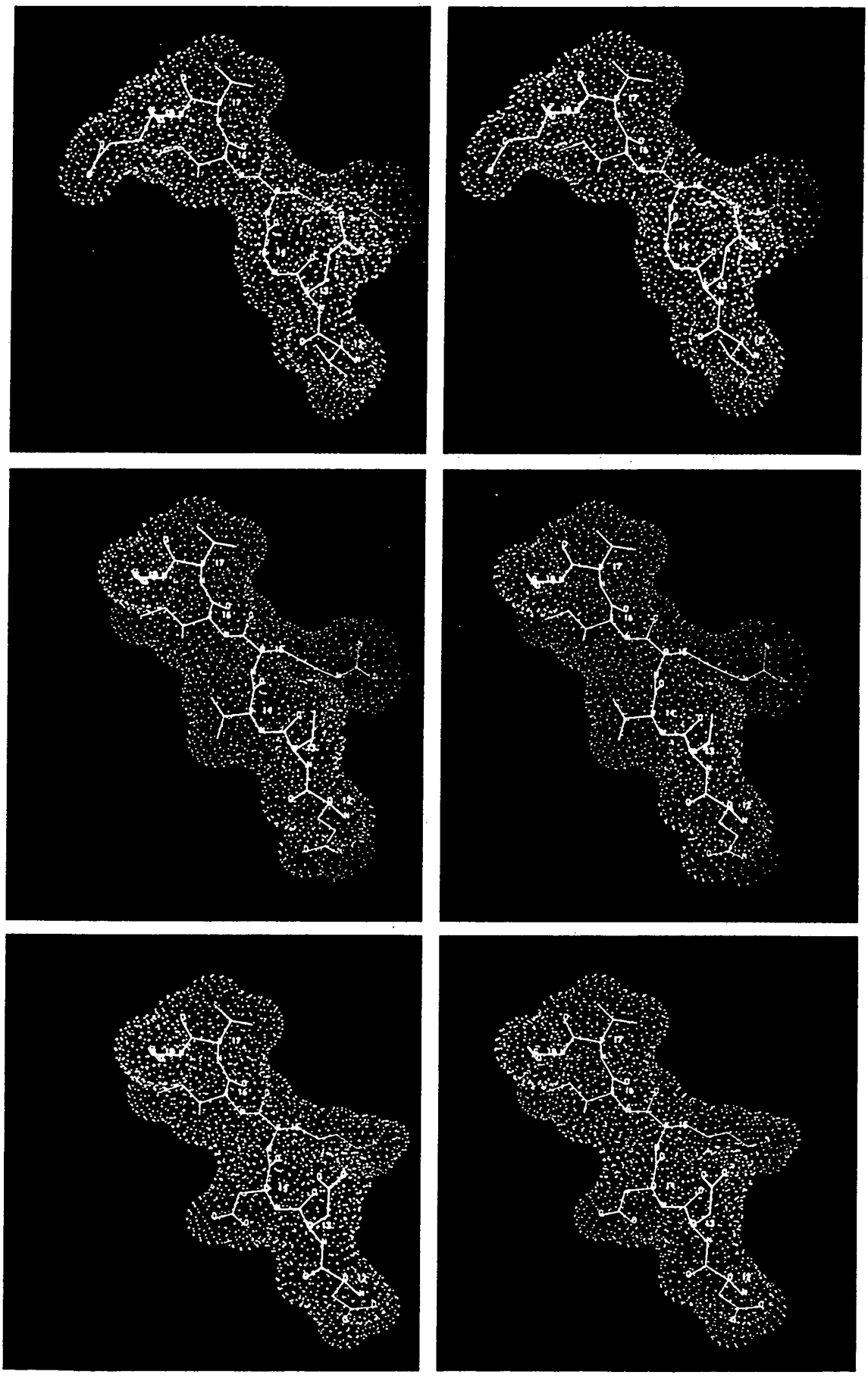

TABLE 4. Distribution of Electrostatic Potential on the Surface Map of the Substrate

Percent of surface area between	Prothrombin	Factor X	Trypsinogen
< −15 kcal/mole	76.6%	0.0%	95.7%
−15 to −5 kcal/mole	8.1	0.0	1.4
−5 to 5 kcal/mole	5.1	0.0	1.8
5 to 15 kcal/mole	3.1	3.5	1.1
> 15 kcal/mole	7.1	96.5	0.0
Average potential (kcal/mole)	−39.5	41.3	−88.0
Total charge on substrate peptide	−1	+1	−2

it would be useful to compare the different complexes systematically to better understand the nature of their specificity.

Langridge and co-workers[23] have introduced a surface map representation of the electrostatic potential of these molecules as a way of depicting the charge-charge interactions between molecules in a complex. When this method is applied to the substrate peptides (FIG. 5) the differences between the various substrates are highlighted. TABLE 4 gives a breakdown of the per-

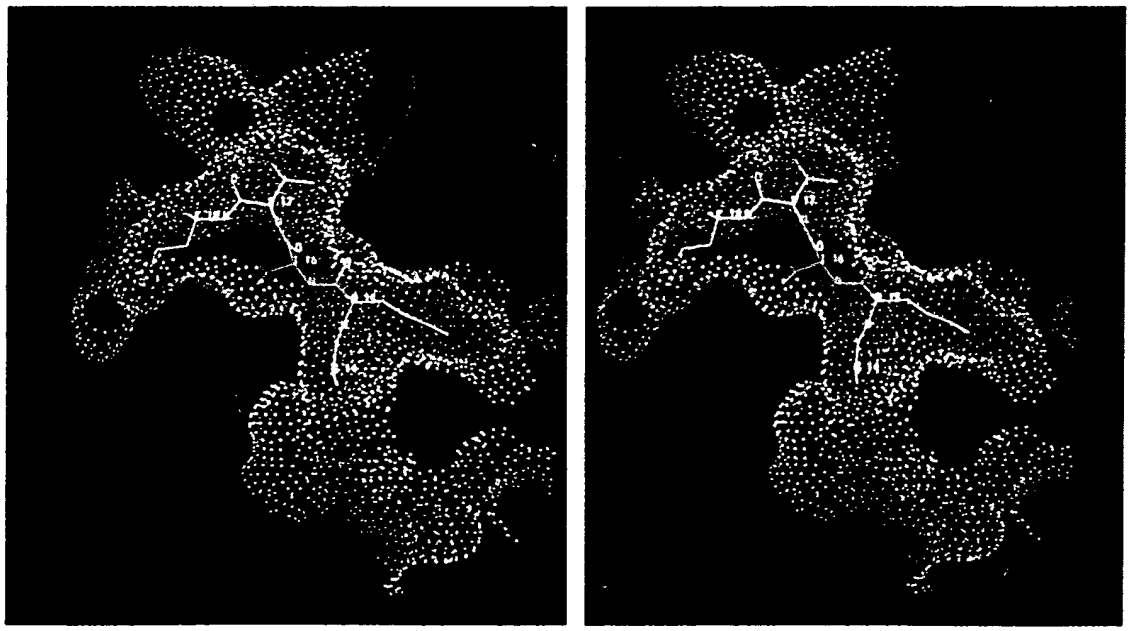

FIGURE 6. A slice through the electrostatic potential surface maps for prothrombin bound to the factor Xa enzyme. The positive surface (*blue*) on the right (labeled R 15) is part of the substrate surface at Arg P_1. It is enveloped by a negative (*red*) surface which is due to the proximity to Asp 189 of factor Xa. On the left, the negative surface is that of Glu P_3' (labeled E 18). Adjacent to it is the positive surface of the enzyme due to Lys 62. Nevertheless, it is very difficult to distinguish which surface comes from the substrate and which from the enzyme, without using different colors for the two surfaces (rather than colors for the electrostatic potential).

cent surface area on each substrate which is positive and negative in order to provide a quantitative indication of the differences between these peptides.

When the electrostatic potential maps are examined in detail, it emerges that the charged side chains, Arg, Lys, Glu, and Asp, dominate the potential function. Therefore, the map for the prothrombin peptide is positive around the Arg P_1 side chain but negative everywhere else because of the two carboxylates of Glu P_3 and P_3' which give the peptide a net charge of -1 (TABLE 4). Similarly, the map for the factor X peptide is almost entirely positive because the only charged species, the positive Arg at P_1, is "felt" everywhere on the surface of the peptide. In the same way, the trypsinogen peptide with a net charge of -2 due to the three negative Asp residues at P_2 to P_4 overwhelms the effect of the positive Lys at P_1 except in the immediate environment of the ε-amino group.

Variation of Substrate-Peptide Electrostatic Potential by Environment

In order to better understand the implications of the electrostatic potential functions, as represented by these color-coded surface maps, to the specificity of the substrate-enzyme interaction, the surface maps of several of the enzymes were examined in the complex with the respective substrate. For this study, the surface maps of the complex of prothrombin with its activating enzyme factor Xa (FIG. 2) are shown in FIG. 6. Unfortunately, the superposition of the two color-coded surfaces makes them difficult to distinguish; consequently, only a thin section can be shown here.[23] It can be seen that, for example, the electrostatic potential on the enzyme surface is opposite to that on the substrate in the region of the salt bridge interactions described above, as shown for Arg P_1-Asp 189 and for Glu P_3'-Lys 62.

Does this mean that the enzyme provides the correct countercharges to balance the charged species on the substrate peptide and yield potential values close to 0? To test this, the electrostatic potential map at the surface of the prothrombin substrate peptide was recomputed, but with the three countercharge residues, Asp 189, Arg 143, and Lys 62 included to give a total charge of zero for the substrate + countercharges system. The resulting map is shown in FIG. 7B and summarized in TABLE 5. The charge distribution is now strikingly different than for the substrate peptide by itself (FIG. 7A). The first point to stress is that very little of the surface area, 13%, is close to 0 potential. Thus, the countercharges are not neutralizing the charge on the substrate but are changing the electrostatic potential distribution considerably. The surface about P_1' to P_3' changes from negative to positive because of its proximity to Lys 62, Arg P_1, and Arg 143. Similarly, the tip of the surface for Arg P_1 becomes negative because of the close approach of Asp 189. Some of the molecular surface does reflect the nature of the atoms that are immediately nearby, rather than longer range charge effects, as for example, the side chain of Ile P_4 is close to zero whereas its carbonyl oxygen is negative.

Having seen the effect of including just the countercharges, how is the electrostatic potential of substrate peptide affected by being placed in the

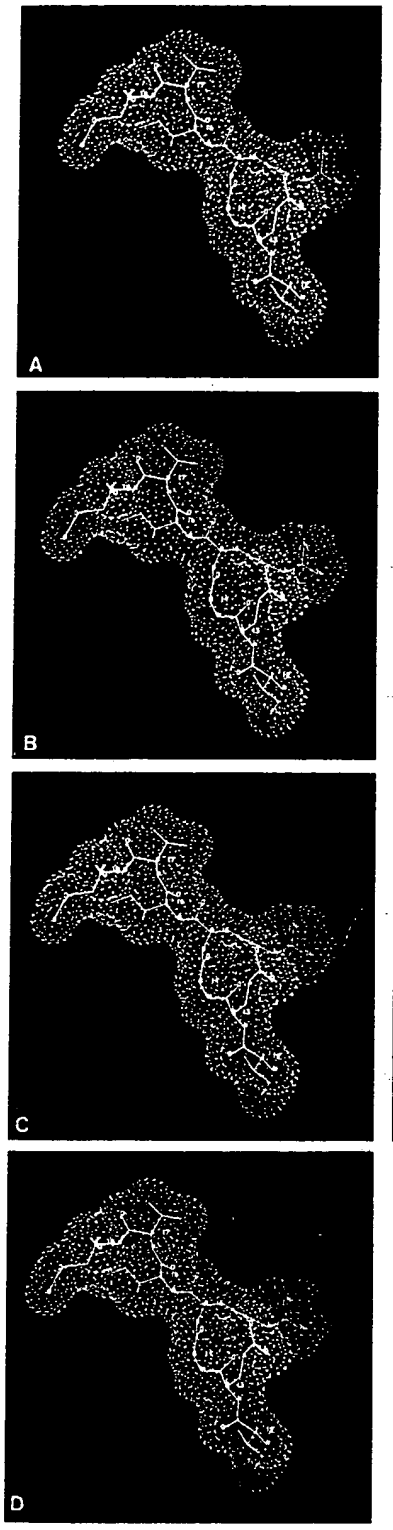

FIGURE 7

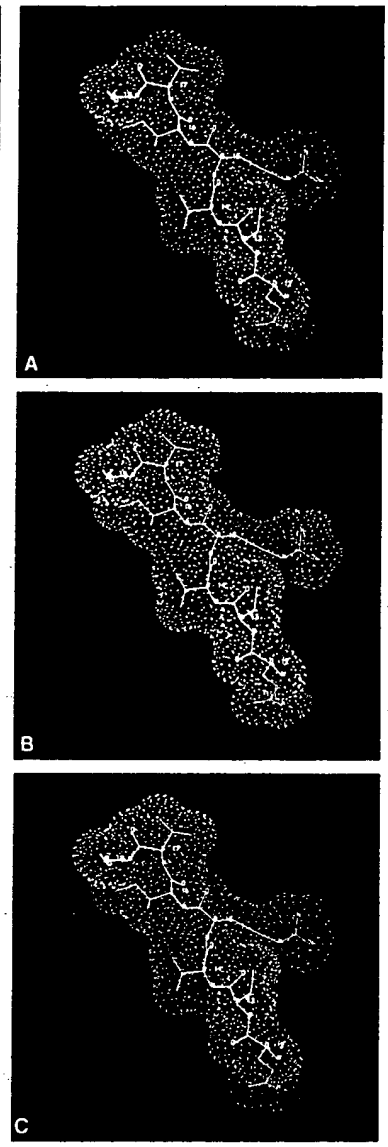

FIGURE 9

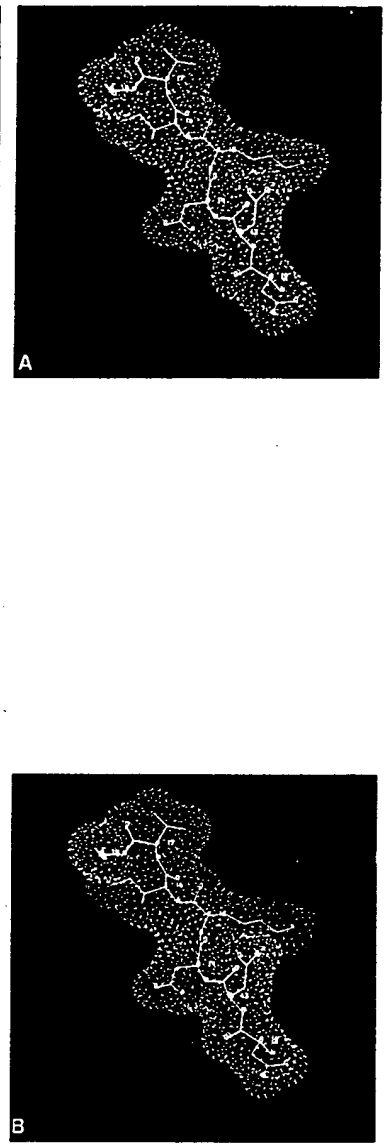

FIGURE 10

TABLE 5. Distribution of Electrostatic Potential for the Prothrombin Substrate Peptide

Percent of surface area between	Prothrombin	Prothrombin + 3 countercharges	Prothrombin + factor Xa	Prothrombin + factor Xa + Ca^{++}
< −15 kcal/mole	76.6%	24.4%	99.9%	70.1%
−15 to −5 kcal/mole	8.1	11.1	0.1	13.3
−5 to 5 kcal/mole	5.1	13.1	0.0	8.2
5 to 15 kcal/mole	3.1	16.5	0.0	6.5
> 15 kcal/mole	7.1	34.9	0.0	1.9
Average potential (kcal/mole)	−39.5	2.6	−58.4	−27.3
Net charge on system	−1	0	−1	+1

complete environment of the factor Xa enzyme model structure? The net charge of the factor Xa molecule is 0. This gives a total charge of −1 for the enzyme-substrate complex. The result of this calculation was a complete surprise; FIG. 7C and TABLE 5 present the results. The entire map is negative with an average electrostatic potential of −58 kcal/mole. Thus, the peptide in the active site of the enzyme must lie in a highly negative environment

FIGURE 7. Electrostatic potential surface maps for the substrate prothrombin activation peptide (see TABLE 5). **(A)** Potential map for prothrombin peptide by itself with a net charge of −1. **(B)** Map for the prothrombin peptide with the three salt bridge residues: Lys 62, Arg 143, and Asp 189. The net charge on this system is 0. **(C)** Potential surface map for the prothrombin peptide on the factor Xa enzyme. Note that the whole surface is now highly negative with mean potential value of −58 kcal/mole. This is due to the distribution of charged residues on the enzyme (see TABLE 6 and text for discussion). **(D)** Ca^{++} placed 14 Å from the substrate near residues Glu 36 and Glu 37 of factor Xa (see FIG. 8). Note that the positive surface corresponds exactly to the carboxylate side chain of Glu P_3' (labeled E 18) which is itself negative.

FIGURE 9. Electrostatic potential surface maps for the substrate factor X activation peptide (see TABLE 7). **(A)** The map for factor X peptide which has a net charge of +1. The map is dominated by this positive charge of Arg P_1. **(B)** Potential map for the peptide together with Asp 189, the countercharge for Arg P_1 (labeled R 15). The net charge is now 0. A discussion of this distribution appears in the text. **(C)** Map for the factor X peptide together with the model for its activating enzyme, factor IXa. In the factor IXa enzyme, the active site and specificity sites are highly positive, in contrast to factor Xa (FIG.7C), where the environment was highly negative. The potential at the surface of residue Val P_2' (labeled V 17) is negative, due to a concentration of negatively charged residues in the region of the enzyme adjacent to this position (see text).

FIGURE 10. Electrostatic potential surface maps for the substrate trypsinogen activation peptide (see TABLE 8). **(A)** The map for the trypsinogen peptide alone with a charge of −2 due to the Asp residues. The surface is overwhelmingly negative. **(B)** The trypsinogen peptide in the environment of trypsin gives this potential map. The *N*-terminal portion remains negative since there are no countercharges for Asp P_2 and Asp P_3 (labeled D 14 and D 13, respectively) and they point out into solvent. There is a countercharge for Asp P_4 (labeled D 12) which is Lys 224, and thus some positive surface appears at this residue. The *C*-terminal portion is apparently dominated by the large positive net charge on the trypsin molecule of +10. Note that the transition from negative to positive occurs right at the point of the cleaved peptide bond.

TABLE 6. Net Charge Distribution on Enzyme in Shells about the Substrate

Substrate	Prothrombin		Factor X		Trypsinogen	
Total charge	−1		+1		−2	
Enzyme	Factor Xa		Factor IXa		Trypsin	
Total charge	0		+1		+10	
Shell range (Å)	Net charge in shell	Accumulated net charge	Net charge in shell	Accumulated net charge	Net charge in shell	Accumulated net charge
0–3	0	0	−1	−1	0	0
3–5	−1	−1	0	−1	−2	−2
5–7	−1	−2	−1	−2	+1	−1
7–9	0	−2	+3	+1	0	−1
9–12	0	−2	+2	+3	+4	+3
12–15	+1	−1	+1	+4	+3	+6
15–20	+2	+1	+1	+5	+1	+7
20–50	−1	0	−4	+1	+3	+10

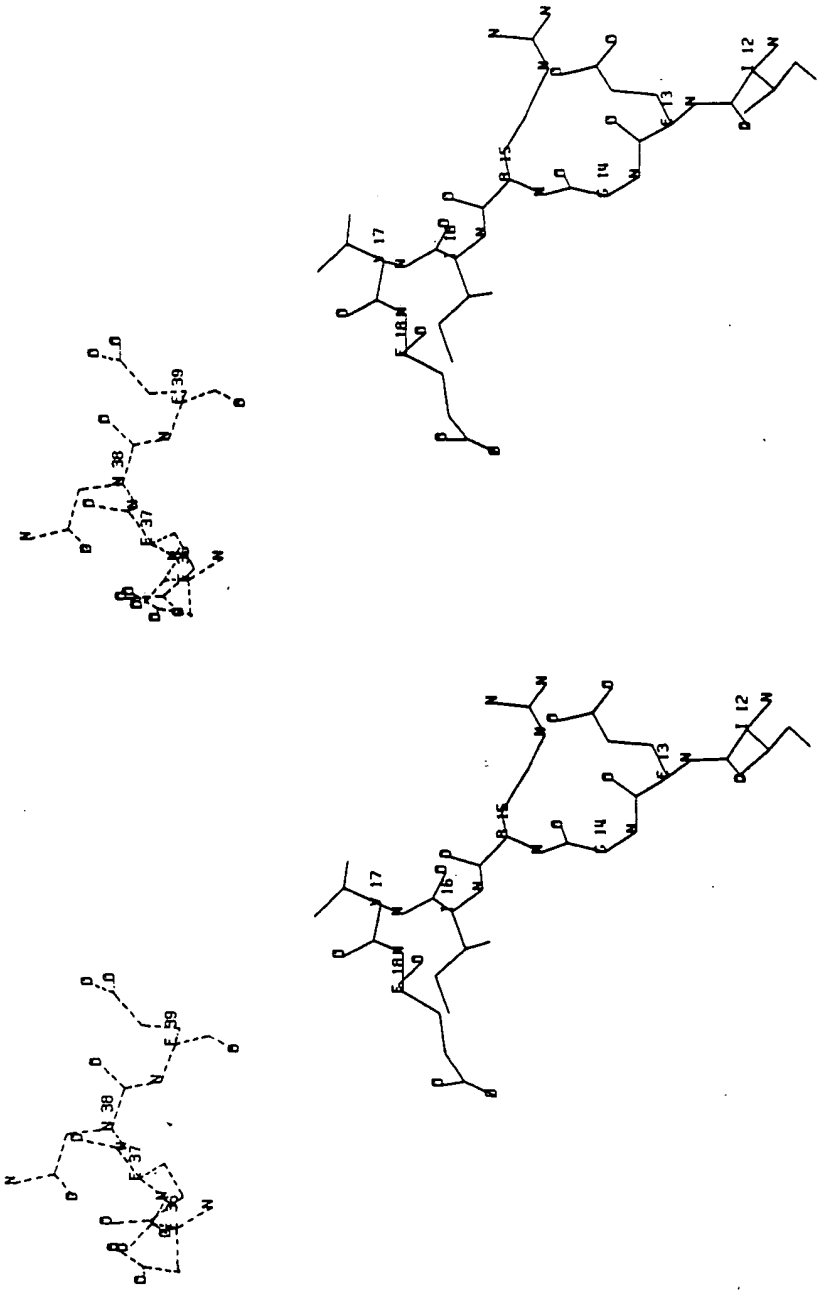

FIGURE 8. The loop of factor Xa containing residues Glu 36, Glu 37 and Glu 39 is shown (*dashed lines*) in relation to the prothrombin substrate (*solid lines*). The position of the added Ca^{++} ion is marked by a "#" in upper left of the figure near the carboxylates of Glu 36 and Glu 37. The nearest distance from the Ca^{++} ion to Glu P_3' (labeled E 18) of prothrombin is 14.4 Å.

even though, as noted above, the net charge on the enzyme is 0. In order to determine why this is the case, the net charge on the enzyme was calculated in shells around the substrate, as shown in TABLE 6. As can be seen, factor Xa has an excess of negative charges within 7 Å of the substrate. There are shells with an excess of positive charge also, but these lie much farther away, from 12 to 20 Å, and thus by Coulomb's law, have a much smaller effect on the substrate. It is particularly striking that the active-site environment of factor Xa is so negative, when the net charge on the prothrombin peptide is also negative.

Is this highly negative environment of the active site a realistic view of this enzyme-substrate interaction in the factor Xa-prothrombin complex? Clearly, these charge effects could be modified and modulated by the binding of ions from solvent and buffer onto the protein. In particular, there is a loop in factor Xa that contains three Glu residues very close together, at positions 36, 37 and 39 (FIG. 8). Two, and perhaps all three, of these residues may form a Ca^{++} ion binding site. In order to test the effect of the binding of such an ion, a Ca^{++} was bound to the enzyme at this site and given a charge of +2.0. It is important to point out that the closest this ion approaches to the substrate peptide is 14.1 Å at Glu P_3' (see FIG. 8).

The electrostatic potential map of the prothrombin activation peptide was recalculated in the environment of the factor Xa enzyme model with the one Ca^{++} ion bound, as described. The resulting map is shown in FIG. 7D and is summarized in TABLE 5. There is a significant influence of this single Ca^{++} ion on the electrostatic potential in the environment of Glu P_3' and its surrounding, which has changed from significantly negative to positive. Thus, it is clear that a proper treatment of the electrostatic environment of the substrate must take into account solvent and buffer ion interaction with the enzyme even at sites somewhat removed from the immediate region of the active and specificity sites of the enzyme.

Similar calculations were performed on the electrostatic potential map of the factor X activation peptide substrate bound to its activating enzyme factor IXa (see FIG. 9 and TABLE 7). As previously discussed, the substrate peptide appears to be entirely positive with a mean electrostatic potential value of 41 kcal/mole (Fig. 9A) due to the Arg at the P_1 site. When the countercharge for this Arg, Asp 189 from the protein, is included in the calculation of the electrostatic potential, the result is a significant difference in the potential function (Fig. 9B and TABLE 7) which now has an average potential value of 8.4 kcal/mole. Very little of the surface is actually negative, mostly in the immediate region of Asp 189. There is a positive band that is largely due to the influence of the Arg P_1 guanidinium group. Moving further away from Arg P_1, there is a belt of approximately neutral or slightly positive or negative potential that is largely the result of the carbonyl oxygens of residues P_2 and P_4 and the Gln P_4 side chain. Finally the edge of the molecule which points mostly away from the activating enzyme (not easily seen in this view) is again positive. This appears to be due to a series of main chain amino hydrogens of residues P_2, P_1 and P_1' all pointing in a similar direction (see FIG. 3). Once again the addition of a countercharge, which gives a formally

TABLE 7. Distribution of Electrostatic Potential for the Factor X Substrate Peptide

Percent surface area between	Factor X	Factor X + 1 countercharge	Factor X + factor IXa
< −15 kcal/mole	0.0%	4.3%	0.0%
−15 to −5 kcal/mole	0.0	7.7	1.9
−5 to 5 kcal/mole	0.0	24.7	1.9
5 to 15 kcal/mole	3.5	36.1	3.0
> 15 kcal/mole	96.5	27.3	93.2
Average potential (kcal/mole)	41.3	8.4	64.9
Net charge on system	1	0	2

net uncharged system, results in a distribution of very significant positive and negative potentials rather than a close to zero value.

Next, the electrostatic potential map was recalculated for the factor X peptide in the environment of the factor IXa model structure (FIG. 9C and TABLE 7). The resulting map is different from both the isolated peptide and the peptide-countercharge distribution. The potential map is almost entirely positive. This is not the result of the overall +1 charge on the factor IXa molecule (versus −1 for factor Xa, see TABLE 6) because the addition of +2 to factor Xa (FIG. 7D) had only a limited effect on the negative environment even though the total net charge of factor Xa + Ca^{++} is +1. Rather, it is once again the result of the charge distribution on factor IXa (TABLE 6). Whereas the immediate region about the substrate from 0–7 Å tends to be negative, a larger net positive charge from 7–15 Å seems to overwhelm the negative group and gives the active site a positive character.

Interestingly, despite the overall positive environment, the surface about the side chain and carbonyl oxygen of Val P_2' is quite negative. This is due to a concentration of negative charges and a paucity of positive charges in this region of the factor IXa enzyme, in particular, Glu 36 and Glu 74 but also Glu residues at positions 70, 78, and 80.

TABLE 8. Distribution of Electrostatic Potential for the Trypsinogen Substrate Peptide

Percent surface area between	Trypsinogen	Trypsinogen + trypsin
< −15 kcal/mole	95.7%	30.1%
−15 to −5 kcal/mole	1.4	9.4
−5 to 5 kcal/mole	1.8	5.9
5 to 15 kcal/mole	1.1	3.4
> 15 kcal/mole	0.0	51.2
Average potential (kcal/mole)	−88.0	21.3
Net charge on system	−2	+8

The average potential of 65 kcal/mole for factor X on the factor IXa enzyme indicates that the factor X substrate peptide, which itself has a net positive charge, resides in a highly positive environment in factor IXa. This is the exact parallel of the situation in the enzyme factor Xa, which was a negative substrate peptide in a highly negative environment (see FIG. 7C). Thus, in each case, the enzyme active site carries a potential that has the same sign as the potential on the substrate peptide. One might speculate that this may represent a mechanism whereby the enzyme prevents too strong binding of the substrate—which would result in diminished turnover.

Similarly dramatic differences can be observed in the electrostatic potential function of the trypsinogen activation peptide on the enzyme trypsin (FIG. 10 and TABLE 8). In the isolated substrate peptide, as noted above, the three Asp residues at P_2 through P_4 dominate the potential function, virtually overwhelming the positive charge at Lys P_1. Because countercharges for Asp P_2, Asp P_3 and Asp P_4 (except perhaps Lys 224 for the latter, see FIG. 4), do not occur on trypsin, the potential map for the substrate peptide together with enzyme countercharges cannot be computed. However, the addition of the enzyme which has a net +10 charge has a remarkable effect on the potential function. The *N*-terminal half of the substrate peptide remains negative, due presumably to the aspartate residues. Only Asp P_4 has a possible countercharge on the enzyme in residue Lys 224, hence the positive and zero region at the surface of residue P_4. The *C*-terminal half of the molecule does apparently "feel" the highly positive nature of the molecule. The complete substrate is not dominated by the +10 charge of trypsin because most of the positive side chains of trypsin are distant from the substrate (see TABLE 6). It is especially clear that both with respect to the Asp residues at positions P_2 through P_4 and to the many excess positive residues on trypsin, all of which are exposed to solvent, the binding of ions from the buffer and solvent should play a crucial role in the proper evaluation of the electrostatics of substrate-enzyme interaction.

Because trypsin is a digestive enzyme, it tends to be less specific than the other serine proteases studied here. Therefore, it probably operates mostly on denatured proteins or highly accessible peptide portions of macromolecules. Thus, it is likely that a substrate peptide such as in trypsinogen will bind to trypsin with Asp residues P_2 and P_3 and perhaps even P_4 (despite the Lys 224 counterion) interacting ultimately with solvent and buffer ions. Hence, the electrostatic effect of these aspartates would be considerably diminished. It will be interesting to examine several other possible substrates of trypsin in order to see how they are affected by the trypsin environment and compare these results with those obtained for the more specific blood clotting factors described above.

ACKNOWLEDGMENTS

I thank Drs. T. J. O'Donnell and Arthur Olson for providing the GRAMPS graphics program and Dr. Michael Connolly for the solvent-exclusion sur-

face program. This study also benefited from graphics programs written by Fred Thomas and Jeff Wakat.

REFERENCES

1. Browne, W. J., A. C. T. North, D. C. Phillips, K. Brew, T. C. Vanaman & R. L. Hill. 1969. J. Mol. Biol. **42**: 65–86.
2. McLachlan, A. D. & D. M. Shotton. 1971. Nature New Biol. **229**: 202–205.
3. Jurasek, L., R. W. Olafson, P. Johnson & L. B. Smillie. 1976. *In* Proteolysis and Physiological Regulation. D. W. Ribbons & K. Brew, Eds.: 93–123. Academic Press. New York.
4. Kretsinger, R. H. 1976. Annu. Rev. Biochem. **45**: 239–266.
5. Greer, J. 1980. Proc. Nat. Acad. Sci. U.S.A. **77**: 3393–3397.
6. Greer, J. 1981a. J. Mol. Biol. **153**: 1027–1042.
7. Greer, J. 1981b. J. Mol. Biol. **153**: 1043–1053.
8. Blundell, T., B. L. Sibanda & L. Pearl. 1983. Nature (London) **304**: 273–275.
9. Birktoft, J. J. & D. M. Blow. 1972. J. Mol. Biol. **68**: 187–240.
10. Tulinsky, A., N. V. Mani, C. N. Morimoto & R. L. Vandlen. 1973. Acta Crystallogr. **29**(B): 1309–1322.
11. Huber, R., D. Kukla, W. Bode, P. Schwager, K. Bartel, J. Diesenhofer & W. Steigemann. 1974. J. Mol. Biol. **89**: 73–101.
12. Stroud, R. M., L. M. Kay & R. E. Dickerson. 1974. J. Mol. Biol. **83**: 185–208.
13. Shotton, D. M. & H. C. Watson. 1970. Nature (London) **225**: 811–816.
14. Sawyer, L., D. M. Shotton, J. W. Campbell, P. L. Wendel, H. Muirhead, H. C. Watson, R. Diamond & R. C. Ladner. 1978. J. Mol. Biol. **118**: 137–208.
15. Titani, K., K. Fujikawa, D. L. Enfield, L. H. Ericsson, K. A. Walsh & H. Neurath. 1975. Proc. Nat. Acad. Sci. U.S.A. **72**: 3082–3086.
16. Katayama, K., L. H. Ericsson, D. L. Enfield, K. A. Walsh, H. Neurath, E. W. Davie & K. Titani. 1979. Proc. Nat. Acad. Sci. U.S.A. **76**: 4990–4994.
17. Schechter, I. & A. Berger. 1967. Biochem. Biophys. Res. Commun. **27**: 157–162.
18. Blow, D. M., J. Janin & R. M. Sweet. 1974. Nature (London) **249**: 54–57.
19. Sweet, R. M., H. T. Wright, J. Janin, C. H. Chothia & D. M. Blow. 1974. Biochemistry **13**: 4212–4228.
20. Segal, D. M., G. H. Cohen, D. R. Davies, J. C. Powers & P. E. Wilcox. 1971a. Cold Spring Harbor Symp. Quant. Biol. **36**: 85–90.
21. Segal, D. M., J. C. Powers, G. H. Cohen, D. R. Davies & P. E. Wilcox. 1971b. Biochemistry **10**: 3728–3738.
22. James, M. N. G., G. D. Brayer, L. T. J. Delbaere, A. R. Sielecki & A. Gertler. 1980. J. Mol. Biol. **139**: 423–438.
23. Langridge, R., T. E. Ferrin, I. D. Kuntz & M. L. Connolly. 1980. Science **211**: 661–666.
24. Connolly, M. L. 1983. Science **221**: 709–713.
25. Blaney, J. M., P. K. Weiner, A. Dearing, P. A. Kollman, E. C. Jorgensen, S. J. Oatley, J. M. Burridge & C. C. F. Blake. 1982. J. Am. Chem. Soc. **104**: 6424–6434.
26. Kollman, P., P. Weiner & A. Dearing. 1981. Ann. N.Y. Acad. Sci. **367**: 250–268.
27. Gelin, B. & M. Karplus. 1979. Biochemistry **18**: 1256–1268.
28. Warshel, A. & M. Levitt. 1976. J. Mol. Biol. **103**: 227–249.
29. Hartley, B. S. 1970. Phil. Trans. Roy. Soc. (series B) **257**: 77–87.